A R T I C L E

Dynamic blending and assimilation in Catalan lingual fricative sequences. An ultrasound and acoustic study

Daniel Recasens 🆔

Universitat Autònoma de Barcelona
Email: daniel.recasens@uab.es

Abstract

Ultrasound and center of gravity frequency data for the sequences /ʃ#s/ and /s#ʃ/ produced by Central Catalan speakers reveal that the former sequence is implemented through continuous articulatory and spectral trajectories which, depending on speaker, may be: intermediate between /ʃ/ and /s/ all throughout, thus supporting a dynamic blending mechanism; /ʃ/-like at onset and intermediate between the two fricatives at offset, which is indicative of C1-to-C2 carryover coarticulation. The sequence /s#ʃ/, on the other hand, undergoes regressive assimilation into [ʃ(ʃ)] according to the acoustic signal but less clearly so in the light of the articulatory data. This discrepancy appears to be due to the fact that, while C1=/s/ assimilates indeed to C2=/ʃ/ at constriction location, coarticulation-induced changes in tongue body configuration behind the primary articulator may occur as long as they do not jeopardize the front-cavity dependent frequency characteristics of the [ʃ] frication noise. Differences in articulatory complexity between /ʃ#s/ and /s#ʃ/ appear to result from the production mechanisms involved, i.e., tongue dorsum raising behind the /s/ constriction for /s#ʃ/ and tongue body repositioning for /ʃ#s/. In agreement with this interpretation, /ʃ#s/ but not /s#ʃ/ turned out to be longer than /s#s/ and /ʃ#ʃ/.

Keywords: dynamic blending; assimilation; fricative sequences; ultrasound; Catalan

1 Introduction

One of the most innovative contributions of Articulatory Phonology to our knowledge of speech production mechanisms has been the concept 'gestural blending' (Browman & Goldstein 1989, 1990, 1992). According to a first formulation, gestural blending is predicted to take place in sequences of two phonetic segments produced with articulatory gestures implemented through the same articulator or contiguous articulators (e.g., either the tongue dorsum or the tongue blade, and the tongue predorsum) and results in a single phonetic realization whose closure or constriction location is intermediate between those for the two original segments. This blending type is exemplified by front velars, i.e., velar consonants followed by a front vowel or palatal glide, which are articulated at the back of the hard palate exclusively or at an articulatory zone embracing the postpalatal and front velar zones (Frisch & Wodzinski 2016). An analogous blending strategy may also yield

laminopredorso-alveoloprepalatal articulations out of sequences of consonants produced with the semi-independent tongue blade and tongue dorsum articulators, as exemplified by a [c]-like realization in the case of the sequences /tk/ and /kt/ (Catalan data reported in Recasens et al. 1993) and by the outcomes [ɲ] and [ʃ] of /nj/ and /sj/, which may occur in fast speech productions of *onion* and *bless you* in English (Zsiga 1995). We have shown elsewhere that there is another gestural blending mechanism involving the addition or summation of the closure or central contact areas of two consecutive consonants. Electropalatographic (EPG) data for Catalan reveal that this blending type applies in sequences of heterosyllabic (dento)alveolar and alveolopalatal stops, nasals and laterals such as /nʎ/, /ɲn/, /ʎn/ and /ɲt/ (where /ʎ/ and /ɲ/ are alveolopalatal), while consonants subject to higher manner requirements such as front lingual fricatives and rhotics fail to take part in this blending process (Recasens 2006).

Two more remarks about the types of gestural blending just mentioned are in order.

On the one hand, the corresponding articulatory blended end product may be highly stable or else continuously changing and thus dynamic over the time domain. The first option applies to sequences of (dento)alveolar and alveolopalatal consonants (and also to front velars); thus, once the closure location for the blended articulation emerges from sequences like /ʎn/ and /ɲt/, it stays unmodified until the C2 release. The dynamic option has been reported to occur for /sj/ and thus a sequence of two consonants i.e., a lingual fricative and a palatal glide, which are both implemented through a temporally variable constriction. Indeed, acoustic centroid and electropalatographic (EPG) data for the sequence /s#j/ in American English, as in *bless you*, show a realization which changes continuously from more /s/-like at onset to more /ʃ/-like at later points in time (Zsiga 1995), and an analogous dynamic articulatory pattern is likely to be at work in instances of the blended outcome [ɲ] of /nj/. Once these two blended outcomes are phonologized (as it has occurred historically in the Romance languages), they become allophones of a single phoneme. On the other hand, and in contrast with regressive assimilatory processes, whenever operating in two-consonant sequences such as those just mentioned, gestural blending is often implemented not until about the midpoint of the temporal portion allocated to C1. This condition appears to hold for instances of static blending (and thus in sequences like /ʎn/ and /ɲt/) as well as for dynamic blending cases (and thus /sj/).

It may be objected that the term 'dynamic blending' is not appropriate for describing the sort of mutual adaptation process which occurs during the production of segmental sequences like /s#j/. Zsiga (1995) found both for the synchronized articulatory and acoustic data that the /s/ palatalization characteristic induced by the palatal glide proceeded gradiently and variably, i.e., the linguopalatal contact pattern and the spectral centroid evolved progressively from more /s/-like to more /j/-like throughout the consonantal sequence (gradience) and speakers could differ as to whether /s/ was approached more or less at onset and /j/ at offset (variability). We believe that the term 'dynamic blending' captures this series of events better than the term 'coarticulation' since the latter implies that segments achieve essentially their target while this is rarely the case for neither /s/ or /j/ in the case of the sequence /s#j/ under consideration.

This paper deals with another case of dynamic gestural blending already reported in previous studies. EPG and spectral center of gravity (COG) data for /s#ʃ/ and /ʃ#s/ and the control sequences /s#s/ and /ʃ#ʃ/ in Central Catalan reported in Recasens & Mira (2013) showed different adaptation strategies for /s/ followed by /ʃ/ (as in *dos xais* 'two lambs') and /ʃ/ followed by /s/ (as in *peix salat* 'salted fish'). In agreement with impressionistic descriptions, the linguopalatal contact data revealed that the sequence /s#ʃ/ is realized as a palatoalveolar fricative all throughout independently of speaker and other factors such as stress placement and speaking rate and consequently may be said to undergo regressive assimilation. The sequence /ʃ#s/, on the other hand, was implemented through a blending

mechanism which resulted into a phonetic realization showing an intermediate lingual configuration between those for /s/ and /ʃ/, which could be considered to be appropriate for the palatalized alveolar fricative [sʲ]. What turned out to be special about the /ʃ#s/ case was that the intermediate lingual configuration in question proceeded from more /ʃ/-like at onset of the two-fricative consonant sequence to more /s/-like at offset and was thus not stable but dynamic. Moreover, the COG data for /s#ʃ/ and /ʃ#s/ were in agreement with the EPG data just described in showing regressive assimilation for /s#ʃ/ and thus stable COG values appropriate for C2 all throughout, and a progressive increase from more [ʃ]-like to more [s]-like COG frequencies for /ʃ#s/. There was also a difference between the articulatory and acoustic data for the latter consonantal sequence, namely, the dynamic trajectory for the EPG signal was closer to /ʃ/ than to /s/ (not only at onset but later in time as well) and the reverse for the COG signal. This difference, which was not referred to explicitly in our previous study and will be commented upon in some detail in the Discussion section of the present paper, ought to be due to the fact that, while the electropalatographic data values were associated with the overall linguopalatal contact pattern (they were obtained using the EPG similarity index method developed by Guzik & Harrington 2007), the COG values of the frication noise depend exclusively on the size of the oral cavity located in front of the lingual constriction (Fant 1960). Interestingly enough, the acoustic durations of /s#ʃ/ and /ʃ#s/ were consistent with the assimilated (/s#ʃ/) and temporally blended (/ʃ#s/) realizations in that, while there was no significant difference in duration between /s#ʃ/ and the controls /s#s/ and /ʃ#ʃ/, the frication noise was clearly longer for /ʃ#s/ than for the three other sequences. It should be stated in this respect that, while, as it occurs with any sequence composed of two equal fricatives, rhotics or approximants separated by a word boundary in spoken Catalan and thus consonant sequences like /s#s/, /ʃ#ʃ/, /r#r/ and /j#j/, the long assimilated outcome [ʃː] of /s#ʃ/ is often simplified into [ʃ] (*dos xais* is generally realized as [do ˈʃajs]), this shortening mechanism is expected to operate on /ʃ#s/ only if exhibiting a temporally stable blending trajectory, which, as pointed out above, does not appear to be a feasible option. In other words, one would expect the long fricative outcome of /ʃ#s/ to shorten if showing a [sʲː]-like realization throughout the entire consonantal sequence but less so or not at all if, as the actual production data show, a dynamic blending mechanism is at work and consequently articulatory and acoustic changes occur from a more /ʃ/-like to a more /s/-like configuration.

To summarize, the temporally dynamic blended realization of /ʃ#s/ differs from other temporally stable blending types described above yielding front velar stops and alveolopalatal realizations while paralleling other instances of dynamic blending such as that occurring during the production of /s/ followed by /j/. Moreover, in so far as the articulatory configuration for /ʃ#s/ turns out not to be entirely appropriate for either /ʃ/ at onset or /s/ at offset, it can be suggested that two articulatory targets are not separately programmed in this case but that speakers aim at some compromise phonetic realization between the two fricatives instead.

The production asymmetry between /s#ʃ/ and /ʃ#s/ just discussed appears to have some universal validity though languages may differ in important respects with regard to the precise articulatory implementation of the two consonantal sequences in question. According to EPG and spectral data for English, /s#ʃ/ often has a canonical /ʃ/ quality all throughout, while /ʃ#s/ may show either two separate targets which are appropriate for /ʃ/ and /s/ or else an intermediate trajectory proceeding gradually or abruptly from more /ʃ/-like to more /s/-like (Holst & Nolan 1995; Nolan et al. 1996; Pouplier et al. 2011). On the other hand, EPG and acoustic data for German reveal that /s#ʃ/ may not only undergo regressive assimilation but also blending, while /ʃ#s/ is implemented invariably through a more /ʃ/-like than /s/-like blended realization (Pouplier & Hoole 2016). Finally, spectral data for the frication noises of /s#ʃ/ and /ʃ#s/ in French reported by Niebuhr et al. (2011) show

that /ʃ/ prevails over /s/ leftwards and rightwards and therefore that there is regressive assimilation in the case of /s#ʃ/, as in Catalan and English, and progressive assimilation in the case of the sequence /ʃ#s/, which, albeit less often, may also apply in other languages and dialects such as Western Catalan (see in this respect the EPG and acoustic data for this Catalan dialect reported in Recasens & Mira 2013). In sum, data for the languages reviewed so far show regressive assimilation for /s#ʃ/ most of the time and, regarding the sequence /ʃ#s/, either two separate targets (with regularly some carryover coarticulation yielding a more /ʃ/-like realization than expected at offset), a temporally dynamic blended realization which falls somewhere between /s/ and /ʃ/ or else progressive assimilation. This summary leads us to conclude that the articulation of /ʃ#s/ is far more variable and complex (and thus less controlled) than that of /s#ʃ/ and also that there is a robust trend for /ʃ/ to overcome /s/.

In searching for an explanation for the production data for /s#ʃ/ and /ʃ#s/ in Catalan and the other languages just described, it makes sense to argue that /ʃ/ prevails over /s/ clearly for /s#ʃ/ and in specific cases for /ʃ#s/ since a greater involvement of the dorsum of the tongue during the production of /ʃ/ than of /s/ renders the palatoalveolar fricative more constrained articulatorily than the alveolar fricative. In accordance with this difference in degree of articulatory constraint, /ʃ/ has been reported to be less prone to coarticulate with the adjacent vowels than /s/ while exerting larger coarticulatory effects on them (Recasens & Espinosa 2009). This however cannot explain by itself why phonological assimilation is so much more likely to operate on /s#ʃ/ (and thus at the regressive level) than on /ʃ#s/ (and thus at the progressive level). A plausible interpretation of this asymmetrical behavior, which is also found to hold in the case of other sequence pairs like /tk/-/kt/, may be sought in tongue repositioning for /ʃ#s/ but not for /s#ʃ/. On the one hand, the anticipation of /ʃ/ during preceding /s/ involves a single articulatory action during C1: little interarticulatory coupling between the primary tongue tip articulator and the back of the tongue for apicoalveolar /s/ leaves the tongue body quite free to anticipate the tongue dorsum raising/tongue front lowering gesture for following /ʃ/, which results in some constriction retraction cooccurring with an increase in predorsal contact immediately behind the alveolar constriction. On the other hand, more interarticulatory coupling between the primary laminopredorsal articulator and the back of the tongue and thus a stringent tongue body positioning for /ʃ/ renders the transition from /ʃ/ to /s/ less straightforward. In this case the anticipation of the lingual gesture for /s/ requires the execution of two non-complementary actions, i.e., some tongue dorsum lowering followed by the raising of the tongue tip and blade. We believe that this specific articulatory requirement rather than other factors such as, for example, a higher frequency of occurrence of /s/ vs. /ʃ/ in the word-final position accounts for why, when adjacent to /ʃ/, /s/ undergoes regressive rather than progressive assimilation. A similar but not identical explanation has been proposed by Perkell et al. (1979, 2013): once the lingual groove for the initial fricative has been created, the production of /ʃ/ in the sequence /sʃ/ requires just pushing the tip-blade upward and forward and that of /s/ in the sequence /ʃs/ a more precise tongue front placement along the front–back dimension. This explanatory account assumes that there are differences in articulatory control during the formation of the front lingual constrictions for /ʃ/ (longer and shaped more precisely also to ensure a sufficient sublingual cavity volume) and /s/ (shorter and more ballistic). Differences in articulatory complexity between the two sequences of interest appear to be in accordance with the differences in duration reported above: while assimilated /s#ʃ/ exhibits a comparable duration to /s#s/ and /ʃ#ʃ/, non-assimilated /ʃ#s/ is often longer than /s#s/, /ʃ#ʃ/ and /s#ʃ/.

Language-dependent and even speaker-dependent differences in the articulatory and acoustic implementation of the two consonant sequences of interest may be associated with an aspect which has been largely neglected by previous studies, namely, the precise

constriction location for /s/ and thus whether this consonant is laminopredorsal, more anterior and less grooved, or else apical, more retracted and more grooved. In contrast with the former variant, the latter is more /ʃ/-like and, in the same way as the palatoaveolar fricative, exhibits a sublingual cavity albeit smaller than that for the palatoalveolar cognate. While English may exhibit the two /s/ types (Ladefoged & Maddieson 1996: 146; Dart 1998), French and Italian have typically a predorsal /s/ while Catalan, European Portuguese and Castilian Spanish favor the apical type (see Navarro Tomás 1918: 81–82 for Spanish).

   Within this framework, the purpose of the present investigation is to analyze the realization of the two symmetrical sequences /s#ʃ/ and /ʃ#s/ in Central Catalan using a different technique from those used in previous studies, namely, ultrasound, which provides data on front and back lingual configuration over time. Ultrasound should allow ascertaining whether /s#ʃ/ is implemented through complete regressive adaptation and /ʃ#s/ through dynamic blending or other production mechanisms, and if evidence for these production strategies may be detected not only at about constriction location (and thus at the alveolar zone or prepalate) but also at more posterior areas of the vocal tract (and thus at the palatal, velar and pharyngeal regions). In contrast with EPG, which provides tongue-to-palate contact patterns but no actual lingual configuration data, ultrasound allows collecting information about the placement of the tongue dorsum surface at the palatal/velar zones and of the tongue back at the pharyngeal zone. Gathering lingual configuration contours at different regions of the tongue should make it possible to determine the relative timing of specific articulatory events such as anticipatory coarticulation or gestural blending at different portions of the tongue and thus whether those events occur at a specific tongue region before they do at another region. Detecting the two tongue edges with ultrasound is however problematic since the mandible and hyoid bones refract the sound before it reaches the tongue surface thus creating a black region at both margins of the image where the tongue tip and the tongue root are located (Stone 2005). In order to ascertain the precise articulatory strategies used by speakers for the production of /s#ʃ/ and /ʃ#s/, their lingual configurations will be compared with those for the control sequences /s#s/ and /ʃ#ʃ/ at several consecutive points in time. Moreover, in parallel to our previous study Recasens & Mira (2013), the extent to which articulatory changes are correlated with variations in the acoustic spectrum will be explored through inspection of the lingual contour and spectral data gathered at the same temporal points. Duration values for the fricative sequences will also be evaluated in the light of the following hypotheses: the assimilated palatoalveolar outcome of /s#ʃ/ is expected not to differ from /s#s/ and /ʃ#ʃ/ in duration, while /ʃ#s/ should be longer if exhibiting two clearcut targets or if executed through a dynamic blending mechanism.

2 Method

2.1 *Lingual spline data*

Ultrasound and acoustic recordings were carried out of the following meaningful Catalan sentences including /s#ʃ/, /ʃ#s/, /s#s/ and /ʃ#ʃ/ preceded by a mid front vowel or schwa and followed by [e]: *tu compres xeixa* 'you buy candeal wheat' ([əs#ʃe]); *en coneix cent* '(s)he knows one hundred of them' ([ɛʃ#se]); *en compres cent* 'you buy one hundred of them' ([əs#se]); *allí hi neix xeixa* 'candeal wheat grows up there' ([eʃ#ʃe]). Sentence stress always fell on the initial syllable of the second word and thus the CV portion of the VC#CV sequences of interest. The word *xeixa* was chosen because it is the only meaningful word, together with the noun of the letter 'x' (*xeix*), which begins with stressed [ʃe] in Catalan. The sentence material were recorded six times by five Central Catalan speakers, i.e., two men (DR, the paper's author; RO) and three women (ES, JU, IM) between forty to sixty years of age who speak Catalan on a regular basis in their everyday life.

Ultrasound recordings were performed with an Echo Blaster unit type EB128CEXT from TELEMED and a microconvex Echo Blaster 128 CEXT transducer with a 2–4 MHz frequency range and a central curvature of 20 mm. The ultrasound images were acquired using a probe with a 100% of 104° field of view and a frequency of 2 MHz, which was attached to a transducer holder positioned under the subject's chin in an Articulate Instruments Stabilization Headset. The recording sampling rate was fifty-four frames per second yielding one image every 18.5 ms. Image streams were recorded synchronously with the audio signal sampled at 22,050 Hz with an AKG-D70 microphone using the sync hardware provided with the Articulate Assistant Advanced (AAA) software by Articulate Instruments Ltd. Contours of the back of the alveolar zone and hard palate were also recorded by asking speakers to press the tongue against their hard palate. Tongue contours were tracked automatically at all temporal frames along each C#C sequence token for each speaker using AAA and adjusted manually. Data points for all tongue contours were exported into an ASCII-file as *x-y* coordinates with their origin located at the bottom-left corner of the ultrasound image towards the rear of the vocal tract. Acoustic files were also exported in .wav format for taking segmental duration measures.

Several temporal points for measurement were identified on waveform and spectrographic displays. The onset and offset of the fricative sequences of interest were identified at the onset and offset of the frication noise, which coincided with the end of the preceding vowel and the beginning of the following vowel, respectively. Lingual spline data were processed at those two temporal points (referred to as P1 and P5 from now on), as well as at the consonant sequence midpoint (P3) and at the midpoint between the two resulting P1–P3 and P3–P5 periods and therefore at P2 and P4, respectively. Tongue spline data points were converted from Cartesian to polar coordinates by shifting the origin of the ultrasound image to approximately the center of the ultrasound probe which was located at X= 86.7 mm and Y=0 mm (Mielke 2015). SSANOVA smoothed splines consisting of strings of points separated by 0.01 radians with the associated standard errors (SE) were computed across the splines for all tokens of each C#C sequence using the R package gss to find a best fit curve (Davidson 2006). The rightmost and leftmost edges of the smoothed splines were determined by entering into the SSANOVA computation procedure the corresponding mean angle radian values across all tokens of the consonant sequences of interest. Inspection of the SE values for all sequences and speakers revealed the existence of small deviation percentages from the mean (between 0.25% and 0.5% depending mostly on subject) thus meaning that speakers' productions were highly consistent across sequence tokens.

In order to evaluate the tongue configuration data at several tongue regions, the spatial length of the SSANOVA splines displayed in Cartesian coordinates was divided into four portions which correspond to different articulatory zones, namely, alveolar (ALV), palatal (PAL), velar (VEL) and pharyngeal (PHAR), separately for each subject applying the same subdivision procedure as in a previous publication reporting data for other consonant clusters which were recorded in the same session as the fricative sequences under analysis in the present investigation (Recasens & Rodríguez 2017). The criterion for determining the four articulatory zones is as follows: the boundary between the alveolar and the dental zone was identified at an inflection point occurring at the spline front edge during dental /t/ in the sequence /pt/ and that between the alveolar and palatal zones at another inflection point located at the back alveolar area during the trill /r/ in the sequence /pr/ (/r/ is postalveolar in Catalan); the boundary between the palatal and velar zones was placed at the closure location for the velar stop in the sequence /iki/ which according to EPG data is articulated at the postpalatal zone, just in front of the soft palate, in Catalan; the length of the velar zone was taken to be 1.25 and 1.51 times that of the palatal zone in the case of the male and female speakers, respectively, as reported by Fitch & Giedd (1999); finally,

the pharyngeal zone extended between the left edge of the velar zone all the way until the bottom edge of the lingual splines.

Spatial distances between each of the four lingual regions and the origin of the ultrasound field of view were measured at P1 through P5 on the SSANOVA smoothed splines. The corresponding distance values for the velar and palatal zones were obtained by averaging the distances between the five central points at each zone and the origin in order to avoid possible consistency problems which could have arisen if a single distance using the central point had been computed instead. Given that the splines for the C#C sequences subject to analysis could differ in length and thus be somewhat shorter or longer with regard to each other, the distance values for the two extreme zones, alveolar and pharyngeal, were evaluated by averaging the distances between the origin and five points located not at the zone midpoint but at the upper third of the pharyngeal zone and at the leftmost third of the alveolar zone.

2.2 *Spectral center of gravity*

A frequency measure of the frication noise spectrum, i.e., center of gravity (COG), was computed for all tokens of /s#ʃ/, /ʃ#s/, /s#s/ and /ʃ#ʃ/ using a MatLab script written by Leonardo Lancia (see also Recasens & Mira 2013). COG was calculated by multiplying each frequency in Hertz by the amplitude in decibels at that frequency, summing the products across all frequencies in the relevant range, and dividing the outcoming value by the sum of all the amplitude values. COG measures reflect the mean central frequencies for the entire spectrum and should be inversely dependent on the dimensions of the cavity in front of the lingual constriction (Cho et al. 2002). This COG computation procedure has been applied instead of the COG of the power spectrum (Forrest et al. 1988), the difference between the two methods being that the former assigns more importance to the main spectral peak than the latter (i.e., the latter method divides the spectrum into two chunks of equal power and emphasizes secondary spectral peaks relative to the main spectral peak while also raising the COG values for fricatives with much high-frequency energy). It should be noted in this respect that the noise spectrum of Catalan apicoalveolar /s/ has a main peak at about 4000–5000 Hz which is definitely lower than that for more anterior, laminodental varieties of /s/ (Recasens 2014). COG measures were obtained from FFT spectra over a frequency range spanning from 1–11 kHz using a 25-ms length Hamming window which was shifted through the fricative sequence in steps of 10 ms. Hence the number of spectral slices varied according to the overall duration of the frication noise. The lower frequency limit of the COG range is intended to avoid artifacts due to possible residues of voicing.

Those COG values occurring at approximately the same time points as the lingual splines, i.e., at P1 through P5 as defined above, were selected for statistical analysis. The COG temporal trajectories in question allowed ascertaining how much the fricative spectrum changed throughout the frication noise and at what points in time did the spectral changes take place, if available.

2.3 *Evaluation of assimilation and blending*

Specific criteria were used for deciding whether the fricative sequences /s#ʃ/ and /ʃ#s/ were implemented through assimilation or gestural blending. The sequences /s#s/ and /ʃ#ʃ/ served as controls for determining which one of the two articulatory adaptation processes was at work. Whenever assimilation applies, C1 should equal C2 (regressive assimilation) or C2 should equal C1 (progressive assimilation), and there ought to be small articulatory and/or acoustic changes over time, if any, throughout the entire fricative sequence. For gestural blending to occur, on the other hand, the lingual configuration and/or frication

noise spectrum ought to fall somewhere between /s/ and /ʃ/ throughout the entire fricative sequence, and either exhibit no obvious changes over time (static blending condition) or a continuous change from a more /ʃ/-like to a more /s/-like target in the /ʃ#s/ case and vice versa for /s#ʃ/ (dynamic blending condition). For the evaluation of the production mechanisms for /s#ʃ/ and /ʃ#s/ it will be assumed that the articulatory and acoustic data at the temporal points P1 and P2 correspond to C1 and those at P4 and P5 correspond to C2, while data occurring at point P3 are not strictly associated with either consonant.

2.4 Statistical analysis

Articulatory distance, COG and consonant sequence duration measures were submitted to linear mixed model (LMM) analysis with subject as a random factor using the mixlm package of R version 3.1.2 (R Developmental Core Team 2014).

Linear mixed model tests were run on the distance values between tongue position and the origin of the ultrasound field of view, and on the COG values, with the following fixed factors: sequence (levels /s#ʃ/, /ʃ#s/, /s#s/, /ʃ#ʃ/) and temporal point (P1, P2, P3, P4, P5) for the articulatory and acoustic data, and also articulatory zone (ALV, PAL, VEL, PHAR) for the articulatory data. Additional LMM tests were performed on the tongue distance and COG data for /s#ʃ/, /s#s/ and /ʃ#ʃ/, on the one hand, and for /ʃ#s/, /s#s/, /ʃ#ʃ/, on the other hand, since the main research topic was to find out whether /s#ʃ/ and /ʃ#s/ differed from the control sequences /s#s/ and /ʃ#ʃ/. Statistical results from the latter LMM tests will only be reported since they did not differ substantially from those obtained from the tests performed on the data for all four clusters. Results will be also provided for an LMM analysis performed on the duration values for /s#ʃ/, /ʃ#s/, /s#s/ and /ʃ#ʃ/.

Least Significance Difference (LSD) post-hoc tests were run on all main effects and significant interactions when available in order to find out whether numerical differences between pairs of levels of a given statistical variable reached significance or not. In view of the large number of tests involved in the LMM analyses, the Benjamini-Hochberg (BH) correction procedure for adjusting the false discovery rate was applied to those comparisons which were of relevance to the present investigation. In all statistical tests the significance level was set at $p < 0.05$.

3 Results

3.1 /ʃ#s/

3.1.1 Articulation

The statistical results for the tongue distance data for /ʃ#s/, /s#s/ and /ʃ#ʃ/ are presented in Table 1. According to the top panel there is a highly significant main effect of zone which turned out to be associated with all possible differences among zones but not of temporal point or sequence, and significant interactions for all two- and three-factor combinations. The failure for the temporal point and sequence factors to reach significance indicates that /ʃ#s/ does not differ from either /s#s/ or /ʃ#ʃ/ throughout the entire fricative sequence and thus, that is is implemented not through regressive or progressive assimilation but possibly through gestural blending. More detailed information about the production mechanism used by speakers for the production of /ʃ#s/ ought to come from results for the post-hoc test run on the significant triple factor interaction provided in the middle and bottom panels of the table.

According to the middle panel, significant differences between /ʃ#s/ and /s#s/ (left) were found to hold essentially at all five temporal points and at the three articulatory zones ALV,

Table 1. Statistical results for the tongue distance data from the origin for /ʃ#s/, /s#s/ and /ʃ#ʃ/. (Top) Main effects and significant interactions at three significance levels (*, p < 0.05; **, p < 0.01; ***, p < 0.001). (Middle and bottom) Results from the post-hoc test performed on the three-factor interaction. Filled cells indicate a significant difference and empty cells a non-significant one. The corresponding estimate, standard error and p values are reported in Table A1 of the Appendix.

Zone	$F_{(3, 128)} = 13.31$***
Temporal point	non-significant
Sequence	non-significant
Zone x temporal point	$F_{(12, 128)} = 9.17$***
Zone x sequence	$F_{(6, 128)} = 12.01$***
Temporal point x sequence	$F_{(8, 128)} = 2.02$*
Zone x temp point x sequence	$F_{(24, 128)} = 14.96$***

	/ʃ#s/ vs. /s#s/					/ʃ#s/ vs. /ʃ#ʃ/				
	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
ALV	▨	▨	▨	▨	▨			▨	▨	▨
PAL	▨	▨	▨	▨	▨			▨	▨	▨
VEL								▨	▨	▨
PHAR	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨

	/ʃ#s/									
	P1 vs. P2	P1 vs. P3	P1 vs. P4	P1 vs. P5	P2 vs. P3	P2 vs. P4	P2 vs. P5	P3 vs. P4	P3 vs. P5	P4 vs. P5
ALV		▨	▨	▨	▨	▨	▨	▨	▨	
PAL		▨	▨	▨	▨	▨	▨	▨	▨	
VEL										
PHAR		▨	▨	▨	▨	▨	▨	▨	▨	

PAL and PHAR. On the other hand, differences between /ʃ#s/ and /ʃ#ʃ/ (right) turned out to be significant at all five temporal points P1 through P5 at PHAR and thus at the tongue body back, and essentially at P3, P4 and P5 and thus at the sequence midpoint and later temporal periods but not towards the onset of the two-fricative sequence at more anterior articulatory zones than the pharyngeal zone. It is thus the case that even though the tongue configuration for /ʃ#s/ differs from that for /s#s/ and /ʃ#ʃ/ at all temporal points, it happens to be somewhat more /ʃ/-like at onset than /s/-like at offset. In agreement with this finding, variations in tongue configuration between consecutive temporal points during the /ʃ#s/ sequence (see bottom panel) were significant for P1 vs. P3/P4/P5 and for P2 vs. P3/P4/P5, as well as for P3 vs. P4/P5, at the three articulatory zones ALV, PAL and PHAR, while those
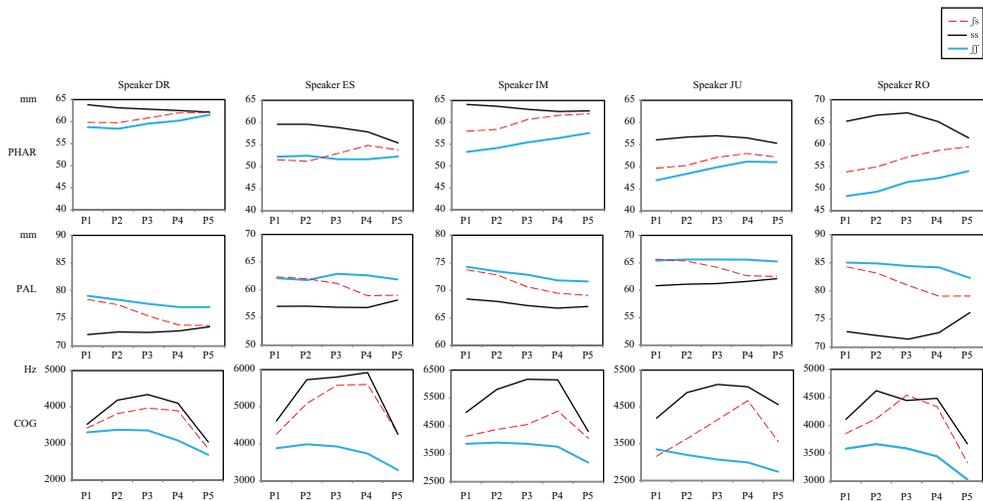
Figure 1. (Colour online) (Top and middle) Distance values between tongue position and the origin of the ultra-sound field of view for /ʃ#s/, /s#s/ and /ʃ#ʃ/ at the temporal points P1, P2, P3, P4 and P5 for the individual subjects. Data correspond to the articulatory zones PHAR (pharynx) and PAL (hard palate). (Bottom) COG trajectories for the same fricative sequences displayed at the same temporal points.

between P1 and P2 and between P4 and P5 did not achieve significance. Consequently, relevant changes in tongue position turned out to occur past P2 and thus towards the middle of the fricative sequence.

In order to get a better feeling for the production strategies for /ʃ#s/ used by the individual speakers, the actual distances in mm between the tongue surface and the origin of the ultrasound field of view at the temporal points P1 through P5 for /ʃ#s/ and for the controls /s#s/ and /ʃ#ʃ/ have been displayed separately for the five individual subjects in Figure 1 (upper and middle row graphs). Distance values are provided for the two articulatory zones PHAR (i.e., distances between the postdorsum and root and the origin) and PAL (i.e., distances between the tongue dorsum and the origin). An increase in tongue surface-to-origin distance at the pharyngeal zone between a given temporal point and the next one means that the tongue back is being retracted while an analogous increase in the tongue-to-origin distance at the palatal zone that the tongue dorsum is being raised.

According to the data presented in Figure 1, tongue distance trajectories over time at PHAR indicate that at time points P1 and P2 the back of the tongue body for /ʃ#s/ (red discontinuous line) remains stable at a location close to or practically identical to /ʃ/ (blue continuous line). Past P1 and P2, the trajectory for /ʃ#s/ proceeds backwards through an intermediate position between the trajectories for /ʃ#ʃ/ and /s#s/ until about point P4. Finally, at the time periods P4 and P5 and thus towards the end of /ʃ#s/, the trajectory in question becomes stable again at a tongue body back position which is not as front as that for /ʃ/ nor as back as that for /s/ or else practically identical to that for /s/. A mirror image occurs at the palatal zone (PAL): /ʃ#s/ starts being very much like /ʃ/, lowers gradually towards /s/ after P2, and stabilizes at P4 somewhere between the tongue dorsum positions for /ʃ/ and /s/ and often somewhat closer to the alveolar than the palatoalveolar fricative. Moreover, a trend for the PHAR and PAL trajectories for /ʃ#s/ to run closer to those for /ʃ#ʃ/ than to the ones for /s#s/ towards the middle of the sequence (and thus around P3) for most speakers may be observed, which is in agreement with the EPG data reported in Recasens & Mira (2013).

A simultaneous consideration of the distance trajectories at PAL and PHAR reveals a tongue position intermediate between those for /ʃ/ and /s/ at sequence onset and offset (whether at PHAR, at PAL or at the two zones) at least in the case of subjects IM and RO, which, in concomitance with the COG data reported in Section 3.1.2, may be considered to correspond to a blending strategy. Moreover, this specific lingual placement may be found mostly at PHAR at onset and at PAL or at PHAR and PAL at offset, thus suggesting that the blending mechanism is implemented at the tongue back before it achieves the tongue dorsum. This sequence of events may also be observed in the overall tongue configuration data for the two subjects IM and RO plotted at time points P1 through P5 in Figure 2b. The other three speakers DR, ES and JU (see also Figures 2a for DR and ES and Figure 2b for JU) exhibit a continuous change from a lingual position appropriate for /ʃ/ at sequence onset to one which is not quite appropriate for /s/ and thus lies closer to /ʃ/ than expected at about the time point P4, which denotes some carryover coarticulation. As referred to above for the lingual trajectories, the tongue configurations for /ʃ#s/ displayed in Figures 2a and 2b also come closer to those for /ʃ/ than to the ones for /s/ for most speakers.

In sum, the lingual profile trajectories for /ʃ#s/ proceed from /ʃ/-like to /s/-like through an articulatory space which is intermediate between the two fricatives and biased towards /ʃ/ at the middle of the sequence, or else exhibit a tongue configuration much like that for /ʃ/ at onset and approaching /s/ at offset. Moreover, this intermediate tongue configuration is available at the pharynx rather than at the palatal zone, i.e., at all or most temporal points at the former articulatory zone and past C1 at the latter.

3.1.2 Spectral center of gravity

Results for the statistical tests run on the COG data for /ʃ#s/, /s#s/ and /ʃ#ʃ/ are presented in Table 2. Statistical results reported in the top panel reveal the presence of significant main effects of temporal point and sequence and a significant temporal point x sequence interaction, all these significant results achieving the $p < 0.001$ level of significance. Post-hoc tests performed on the two main effects yielded significant differences for /s#s/ > /ʃ#s/ > /ʃ#ʃ/ and thus higher values for /s#s/ than for /ʃ#ʃ/ and intermediate ones for /ʃ#s/, which is consistent with the blending account, and for all paired time points except for P3–P4. Results for the post-hoc tests performed on the temporal point x sequence interaction (see the middle and bottom panels of Table 2) differed in some respects from those for the articulatory data reported in Section 3.1.1: there were significant COG differences between /ʃ#s/ and /s#s/ and between /ʃ#s/ and /ʃ#ʃ/ at all five temporal points except for P1 in the case of the latter sequence pair (middle panel); COG differences between pairs of consecutive temporal points turned out to be all significant except for those for P1–P5 and P3–P4 (bottom panel). It is thus the case that, in contrast with the statistical results for the ultrasound data presented in Table 1, the COG frequency values during the temporal periods P1 and P2 allocated to C1=/ʃ/ in the sequence /ʃ#s/ could differ significantly from those for /ʃ/ and /s/, which is consistent with the blending account.

The COG trajectories for /ʃ#s/ for the individual speakers (Figure 1, bottom row graphs) reveal a continuous change from a low frequency value at onset (P1), which is intermediate between /ʃ/ and /s/ (speakers DR, ES, RO) or lies close to or at the same frequency as the COG value for /ʃ/ (speakers IM, JU), to a higher frequency value at P3 and P4 and thus about cluster midpoint and during C2, which approaches the COG frequency for /s/. Past P4, the COG frequency for /ʃ#s/, and those for /s#s/ and /ʃ#ʃ/ as well, lower as the jaw lowers and the constriction size increases in preparation of the following vowel (see Jongman et al. 2000 and Iskarous et al. 2011 for the same anticipatory vowel effect at the offset of the frication period for single fricatives in CV sequences). In so far as it may be implemented by means of a lingual trajectory which runs continuously from more /ʃ/-like to more /s/-like through an
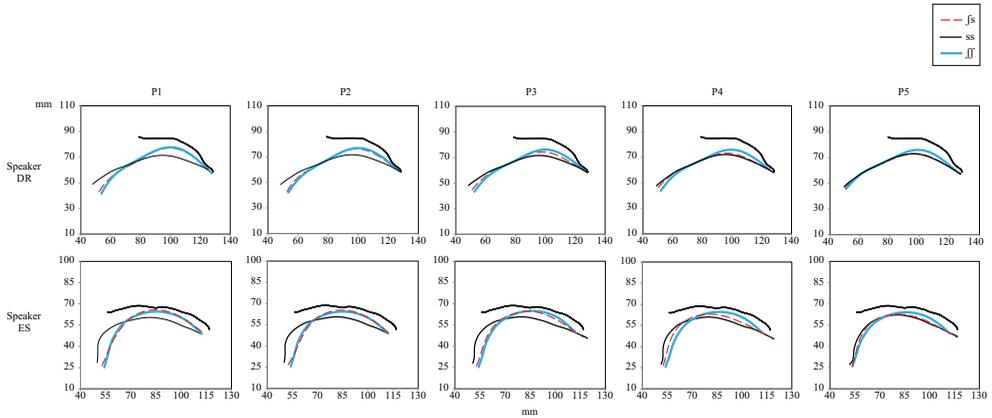
Figure 2a. (Colour online) Lingual configuration data for /ʃ#s/, /s#s/ and /ʃ#ʃ/ at the temporal points P1, P2, P3, P4 and P5 for speakers DR (top) and ES (bottom). The palate surface has been overimposed on each data set. The front of the mouth is on the right side of each figure.
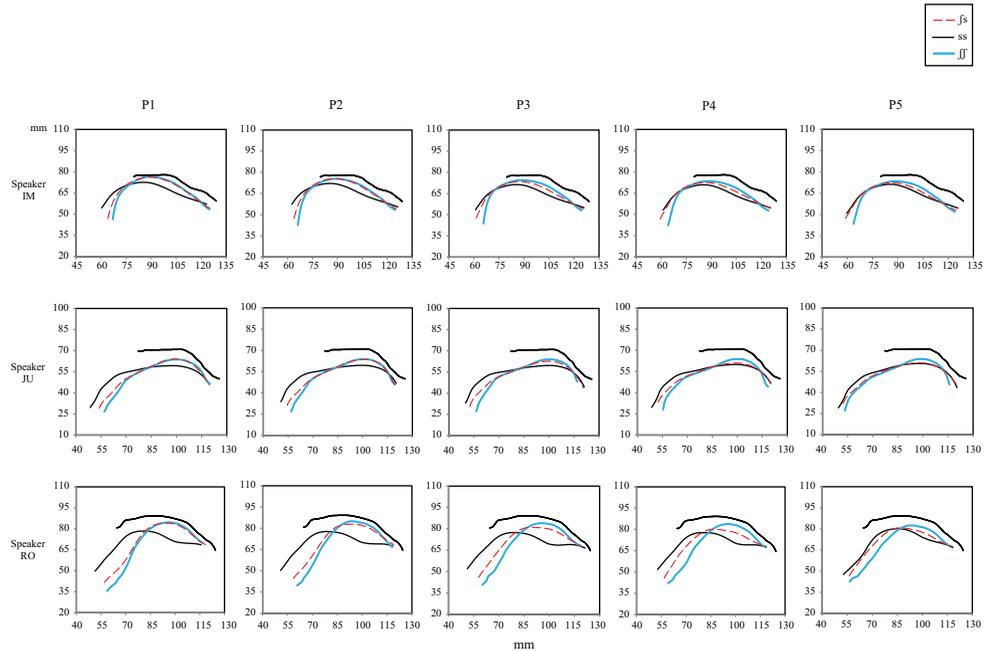
Figure 2b. (Colour online) Lingual configuration data for /ʃ#s/, /s#s/ and /ʃ#ʃ/ at the temporal points P1, P2, P3, P4 and P5 for speakers IM (top), JU (bottom) and RO (bottom). The palate surface contour has been overimposed on each data set. The front of the mouth is on the right side of each figure.

intermediate space between the two fricatives already at sequence onset, the spectral data for /ʃ#s/ appears to be consistent with a gestural blending strategy proceeding dynamically. Moreover, in parallel to results reported in Recasens & Mira (2013), the graphs also reveal a trend for the COG trajectory for /ʃ#s/ to proceed closer to /s#s/ than to /ʃ#ʃ/ for speakers DR, ES and RO (Figure 1). An analysis of the relationship between these COG data and the articulatory data reported in Section 3.1.1 are left for the Discussion section.

Table 2. Statistical results for the COG data for /ʃ#s/, /s#s/ and /ʃ#ʃ/. (Top) Main effects and significant interactions at three significance levels (*, p < 0.05; **, p < 0.01; ***, p < 0.001). (Middle and bottom) Results from the post-hoc test performed on the two-factor interaction. Filled cells indicate a significant difference and empty cells a non-significant one. The corresponding estimate, standard error and p values are reported in Table A2 of the Appendix.

Temporal point	$F_{(4, 375)} = 34.64^{***}$
Sequence	$F_{(2, 375)} = 24.56^{***}$
Temporal point x sequence	$F_{(8, 375)} = 10.31^{***}$

/ʃ#s/ vs. /s#s/					/ʃ#s/ vs. /ʃ#ʃ/				
P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
�	▮	▮	▮	▮	☐	▮	▮	▮	▮

/ʃ#s/									
P1 vs. P2	P1 vs. P3	P1 vs. P4	P1 vs. P5	P2 vs. P3	P2 vs. P4	P2 vs. P5	P3 vs. P4	P3 vs. P5	P4 vs. P5
▮	▮	▮	☐	▮	▮	▮	☐	▮	▮

3.2 /s#ʃ/

3.2.1 Articulation

Results for the statistical analysis performed on the distance values between tongue position and the origin of the ultrasound field of view for /s#ʃ/, /s#s/ and /ʃ#ʃ/ presented in Table 3 show a highly significant main effect of zone (which turned out to be significant for each zone against all the other zones) but not of temporal point or sequence, and several significant interactions among which the three-factor zone x temporal point x sequence interaction. In principle, the absence of a significant sequence effect is not consistent with the hypothesis that /s#ʃ/ should assimilate into [ʃʃ] and thus ought to differ from /s#s/ but not from /ʃ#ʃ/.

Results from the post-hoc tests performed on the triple-factor interaction provided in the middle panel of Table 3 show that the distance values for /s#ʃ/ differ statistically from those for /s#s/ all throughout the fricative sequence at ALV, PAL and PHAR, and from those for /ʃ#ʃ/ at all temporal points at PHAR and at P1 and P2 and thus at the onset of the fricative sequence at ALV and PAL. The finding that /s#ʃ/ differs from /ʃ#ʃ/ mostly during the time allocated to C1 is not in agreement with the regressive assimilation hypothesis; indeed, this difference in tongue position ought not to hold in case that regressive assimilation were at work. According to the bottom panel of Table 3, on the other hand, significant differences between consecutive temporal points occur mostly towards the onset of /s#ʃ/ thus meaning that the tongue is moving substantially during this span of time. In particular, distances at the tongue dorsum (PAL) and tongue body back (PHAR) become significantly different between P1 or P2 and any of the other temporal points, but essentially not for pairs of temporal points involving P3, P4 and P5 exclusively.

Analogously to the ultrasound data for /ʃ#s/ (see Section 3.1.1), in order to make sense of some of these statistical results it is convenient to evaluate them against the actual distance values between the tongue position at PHAR and PAL and the origin of the ultrasound

Table 3. Statistical results for the tongue distance data from the origin for /s#ʃ/, /s#s/ and /ʃ#ʃ/. (Top) Main effects and significant interactions at three significance levels (*, p < 0.05; **, p < 0.01; ***, p < 0.001). (Middle and bottom) Results from the post-hoc test performed on the three-factor interaction. Filled cells indicate a significant difference and empty cells a non-significant one. The corresponding estimate, standard error and p values are reported in Table A3 of the Appendix.

Zone	$F_{(3, 192)} = 13.50$ ***
Temporal point	non-significant
Sequence	non-significant
Zone x temporal point	$F_{(12,192)} = 2.03$*
Zone x sequence	$F_{(6,192)} = 12.08$***
Temporal point x sequence	non-significant
Zone x temp point x sequence	$F_{(24,192)} = 8.91$***

	/s#ʃ/ vs. /s#s/					/s#ʃ/ vs. /ʃ#ʃ/				
	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
ALV	■	■	■	■		■	■			
PAL	■	■	■			■	■	■		
VEL					■	■				
PHAR	■	■	■	■	■	■	■	■	■	■

	/s#ʃ/									
	P1 vs. P2	P1 vs. P3	P1 vs. P4	P1 vs. P5	P2 vs. P3	P2 vs. P4	P2 vs. P5	P3 vs. P4	P3 vs. P5	P4 vs. P5
ALV		■	■						■	
PAL	■	■	■	■	■	■	■			
VEL										
PHAR	■	■	■	■	■	■	■			

field of view at all consecutive points in time (see Figure 3, top and middle row graphs for the data for the individual subjects). At PHAR (top graphs) there is a progressive fronting of the tongue body mostly through the time points P1, P2 and P3 and thus during the first half of the consonant sequence. This tongue body fronting motion proceeds from a more retracted /s/-like position at onset to a more anterior configuration at P4/P5, which either approaches closely or equals that for /ʃ/ (DR, ES, JU) or is intermediate between those for /s/ and /ʃ/ (IM, RO). At PAL (middle graphs), /s#ʃ/ becomes identical or practically so to /ʃ/ but not before the time points P2 or P3, i.e., the tongue dorsum trajectory for /s#ʃ/ starts out somewhere between the trajectories for /ʃ#ʃ/ and /s#s/ and raises until reaching a tongue dorsum position appropriate for /ʃ/ at P3 or P4 (all subjects and most clearly for DR, ES and IM). Taking jointly into consideration the trajectories at PHAR and PAL, it turns out to be the case that practically all speakers show some tongue dorsum raising and tongue back
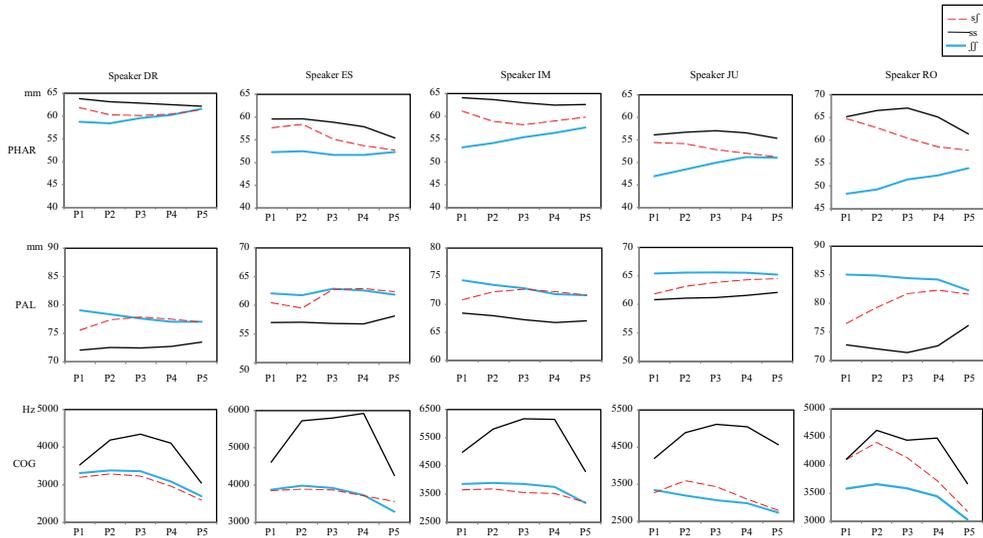
Figure 3. (Colour online) (Top and middle) Distance values between tongue position and the origin of the ultrasound field of view for /s#ʃ/, /s#s/ and /ʃ#ʃ/ at the temporal points P1, P2, P3, P4 and P5 for the individual subjects. Data correspond to the articulatory zones PHAR (pharynx) and PAL (hard palate). (Bottom) COG trajectories for the same fricative sequences displayed at the same temporal points.

fronting during the first half of the sequence /s#ʃ/, which is not consistent with the regressive assimilation hypothesis. This sequence of events may also be observed in the overall tongue configuration data for the individual subjects plotted at time points P1 through P5 in Figures 4a and 4b.

3.2.2 Spectral center of gravity

As to the statistical results for the COG values for /s#ʃ/, /s#s/ and /ʃ#ʃ/, the two main effects (temporal point, sequence) and the temporal point x sequence interaction turned out to be highly significant (Table 4, top panel). Results obtained from the post-hoc tests for those main effects yielded a significant difference between all consecutive time points with a single exception (i.e., the P2–P3 pair) and, more importantly, between /s#ʃ/ and /s#s/ though not between /s#ʃ/ and /ʃ#ʃ/, which is consistent with the assumption that /s#ʃ/ assimilates into [ʃʃ] in Catalan. Also in accordance with the assimilation hypothesis, the pairwise comparisons for the two-factor interaction turned out to be significant for /s#ʃ/ vs. /s#s/ at P1 through P5 but not between /s#ʃ/ and /ʃ#ʃ/ also at all temporal points (Table 4, middle panel). On the other hand, COG frequency differences between consecutive time points did not achieve significance for those pairs involving P1, P2 and P3 but did for pairs involving any of those three time points and P4 or P5 and also in the case of the P4–P5 pair. It was thus the case that COG did not change much during the first half of the consonant sequence but changed a great deal towards its end (Table 4, bottom panel).

Moving now to the COG trajectories over time (see Figure 3, bottom row graphs), it may be seen that the COG frequency values for /s#ʃ/ are practically identical to those for /ʃ#ʃ/ throughout the entire frication noise for the four subjects DR, ES, IM and JU, which is clearly in support of the regressive assimilation account. Moreover, the COG trajectories for all three fricative sequences lower past P3 towards the end of the sequence, which accounts for the significant differences between P1/P2/P3 and P4/P5 referred to above and ought to
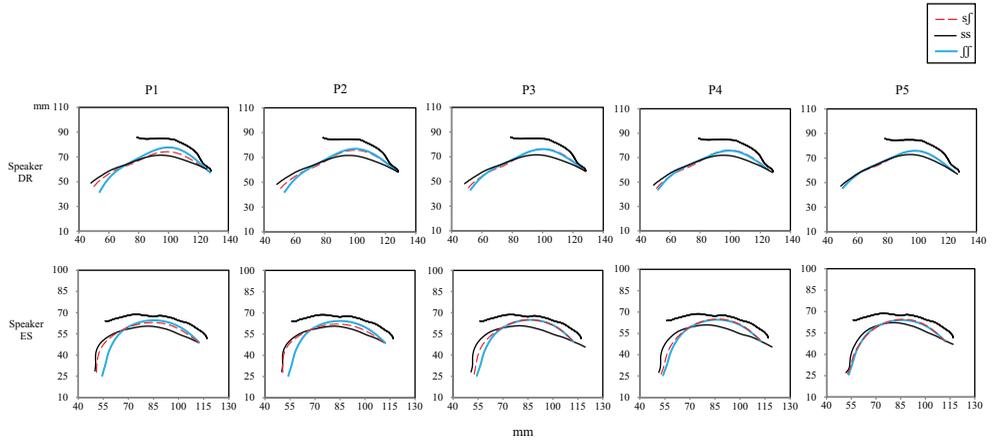
Figure 4a. (Colour online) Lingual configuration data for /s#ʃ/, /s#s/ and /ʃ#ʃ/ at the temporal points P1, P2, P3, P4 and P5 for speakers DR (top) and ES (bottom). The palate surface has been overimposed on each data set. The front of the mouth is on the right side of each figure.
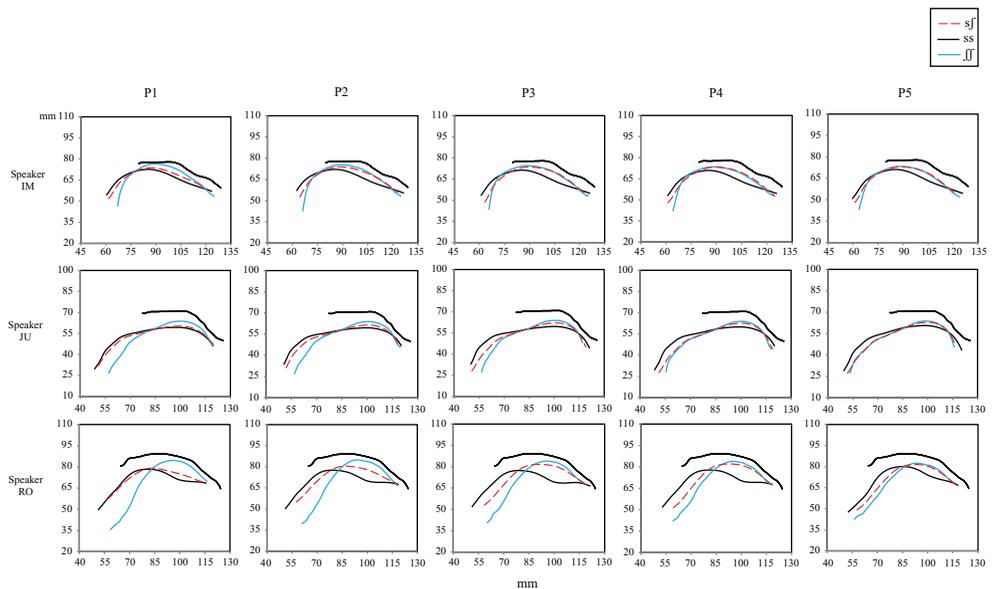
Figure 4b. (Colour online) Lingual configuration data for /s#ʃ/, /s#s/ and /ʃ#ʃ/ at the temporal points P1, P2, P3, P4 and P5 for speakers IM (top), JU (middle) and ES (bottom). The palate surface has been overimposed on each data set. The front of the mouth is on the right side of each figure.

be attributed to the gradual jaw lowering and constriction opening motion occurring at the offset of the frication period in anticipation of the following vowel (see also Section 3.1.2).

Judging from the articulatory and acoustic analysis results presented for /s#ʃ/ in this section and in Section 3.2.1, it appears that the regressive assimilation account is more consistent with the spectral data than with the articulatory data: indeed, while the COG values for this fricative sequence do not differ from those for /ʃ#ʃ/ since its very onset, the

Table 4. Statistical results for the COG data for /s#ʃ/, /s#s/ and /ʃ#ʃ/. (Top) Main effects and significant interactions at three significance levels (*, p < 0.05; **, p < 0.01; ***, p < 0.001). (Middle and bottom) Results from the post-hoc test performed on the two-factor interaction. Filled cells indicate a significant difference and empty cells a non-significant one. The corresponding estimate, standard error and p values are reported in Table A4 of the Appendix.

Temporal point	$F(4, 375) = 53.15^{***}$
Sequence	$F(2, 375) = 21.42^{***}$
Temporal point x sequence	$F(8, 375) = 10.01^{***}$

/s#ʃ/ vs. /s#s/					/s#ʃ/ vs. /ʃ#ʃ/				
P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
▨	▨	▨	▨	▨					

/s#ʃ/									
P1 vs. P2	P1 vs. P3	P1 vs. P4	P1 vs. P5	P2 vs. P3	P2 vs. P4	P2 vs. P5	P3 vs. P4	P3 vs. P5	P4 vs. P5
		▨	▨						

corresponding tongue configurations do not become /ʃ/-like until about the sequence mid-point. A tentative interpretation for this discrepancy will be proposed in Section 4. For the time being it suffices to say that gestural blending is not at work here since /s#ʃ/ becomes the same as /ʃ#ʃ/ whether throughout the entire sequence (COG) or during its second half (articulation).

3.3 Articulatory data summary

In order to draw a cross-speaker comparison between the articulatory and acoustic trajectories for the two asymmetrical sequences /ʃ#s/ and /s#ʃ/, a data normalization procedure has been carried out by setting the value for /s#s/ to 1 and that for /ʃ#ʃ/ to 0 for the PHAR and COG signals, and the reverse for PAL, at all five time points and rescaling the temporal values for the asymmetrical sequences /ʃ#s/ and /s#ʃ/ accordingly. In the three graphs of Figure 5 the normalized PHAR, PAL and COG trajectories for /ʃ#s/ and /s#ʃ/ have been overlaid for comparison.

It may be seen that all trajectories for /ʃ#s/ (continuous line) proceed from a C1-like to a C2-like target, the consonant target being fully achieved only for /ʃ/ in the case of the PAL trajectory. Moreover, a look at the articulatory and acoustic values at sequence midpoint (and thus at P3) shows that changes in tongue position run closer to the /ʃ/ space than to the /s/ space and the reverse for the spectral changes. Regarding the sequence /s#ʃ/ (discontinuous line) full regressive assimilation may be inferred from the COG trajectory in so far as it is essentially identical to that for /ʃ/ all throughout. The PAL distance trajectory (though not the PHAR trajectory, which shows considerable speaker-dependent variability past P3), on the other hand, achieves the lingual target for /ʃ/ at about sequence midpoint but not at time points P1 and P2 during C1. An explanatory hypothesis for why regressive
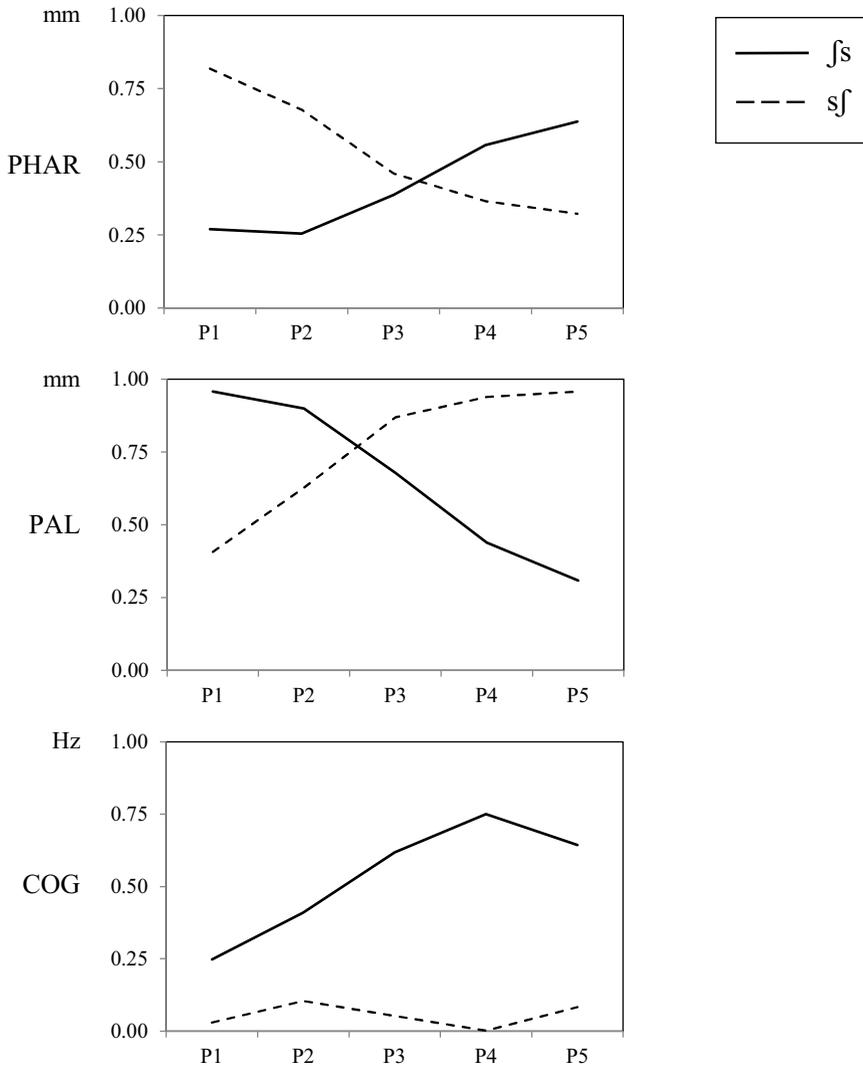
Figure 5. Normalized tongue position and COG values for /ʃ#s/ and /s#ʃ/ at the temporal points P1, P2, P3, P4 and P5 across subjects. Values for /s#s/ are set to 1 and those for /ʃ#ʃ/ to 0 for the PHAR and COG signals, and the reverse for PAL.

assimilation during C1 is apparent in the COG data but not in the articulatory data will be provided in Section 4.

3.4 Duration

The fricative sequence duration values achieved significance according to the LMM test carried out on /ʃ#s/, /s#s/ and /ʃ#ʃ/ (F(2, 75)=7.35, $p < 0.05$) but not to the test performed on /s#ʃ/, /s#s/ and /ʃ#ʃ/. In agreement with data reported in our previous study (Recasens & Mira 2013), /ʃ#s/ turned out to be significantly longer than /s#s/ and /ʃ#ʃ/, which is consistent with the production of the former sequence exhibiting changes in articulatory state
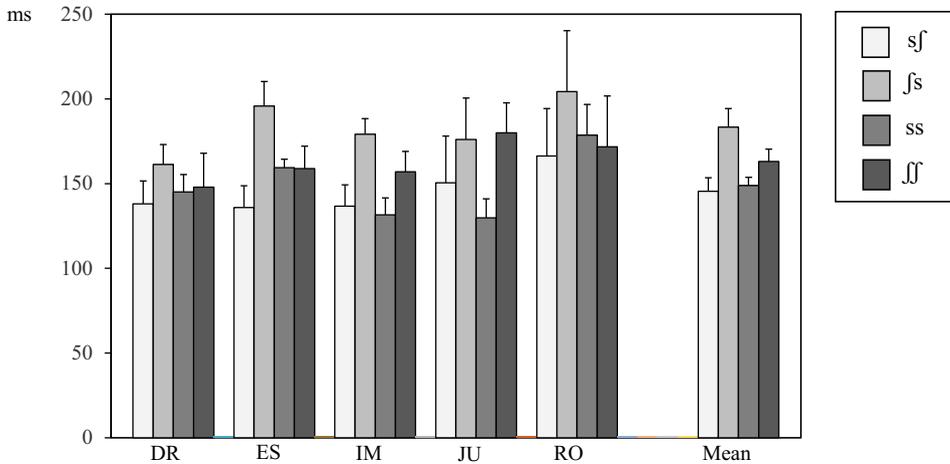
Figure 6. (Colour online) Duration values in ms and standard deviations for /s#ʃ/, /ʃ#s/, /s#s/ and /ʃ#ʃ/ across speakers and for the individual subjects.

over the time domain. The bar graph of Figure 6 shows that all subjects produced indeed a longer frication noise for /ʃ#s/ than for /s#ʃ/, /s#s/ and /ʃ#ʃ/. On the other hand, the fact that /s#ʃ/ had a comparable duration to /s#s/ and /ʃ#ʃ/ suggests that it was articulated with a single target and therefore that the alveolar fricative assimilates to the following palatoalveolar fricative in constriction location in this case.

Judging from differences in sequence duration among speakers, it may very well be that the reason why speaker RO did not assimilate /s/ to following /ʃ/ (see Figures 3 and 4b) is because he produced all four fricative sequences more slowly than the other subjects. In spite of this he showed signs of a blending strategy in the case of the sequence /ʃ#s/ (see Figures 1 and 2b).

4 Discussion

Articulatory data for Catalan reported in the present study reveal that the production of /ʃ#s/ involves a change in tongue position from onset to offset and thus a temporally dynamic trajectory, which appears to be in conformity with the production characteristics of single lingual fricatives in many respects (Iskarous et al. 2011; Reidy 2015). A longer duration for /ʃ#s/ than for the controls /s#s/ and /ʃ#ʃ/ is in line with the dynamic nature of the articulatory trajectory in question.

This articulatory change proceeds through intermediate targets for /ʃ/ and /s/ for some subjects or else from a lingual configuration appropriate for /ʃ/ to a configuration which is intermediate between /ʃ/ and /s/ for other speakers, and tends to be more /ʃ/-like than /s/-like not only at onset but towards the sequence midpoint as well. Towards the onset of the fricative sequence, this intermediate configuration happens to be more apparent at the pharynx than at the palatal zone where the tongue dorsum position for C1 is practically identical to /ʃ/. On the other hand, for most subjects, the COG trajectories proceed from more /ʃ/-like to more /s/-like through intermediate frequencies at all time points, and tend to be more similar to /s/ than to /ʃ/.

It may be argued that, even though the articulatory outcome for /ʃ#s/ is not temporally stable as that for front velar stops or for two-stop consonant sequences involving

the superposition between the closure areas for C1 and C2 (see Introduction), we are facing, at least for some subjects, a blending strategy yielding an articulatory configuration which is intermediate between those for the alveolar and palatoalveolar fricatives. EPG data for /ʃ#s/ in Catalan reported in Recasens & Mira (2013) also show a dynamic realization proceeding continuously from a more /ʃ/-like to a more /s/-like configuration, which may be implemented through a decrease in dorsopalatal contact and/or an increase in lingual constriction fronting along the time domain. Strictly speaking this is not necessarily a two-target movement since the lingual targets for the two consonants of the sequence /ʃ#s/ (i.e., palatoalveolar and alveolar, respectively) may be approached but not fully achieved. The discrepancy between the articulatory and COG data reported in Sections 3.1.1 and 3.1.2 suggests that the tongue front and also the tongue root for some speakers is being fronted towards the /s/ target already at the onset of the fricative sequence while the tongue dorsum lags behind. This tongue tip/blade fronting motion, which cannot be detected in the ultrasound signal (see Introduction), may account for why the COG frequency value increases between P1 and P2 and thus the temporal portion assigned to the first fricative of the consonantal sequence.

The scenario for /s#ʃ/ was more complex than initially expected in so far as a clear assimilatory mechanism was at work for COG but not for the articulatory data while being consistent with the presence of non-significant differences in duration between /s#ʃ/, on the one hand, and /s#s/ and /ʃ#ʃ/, on the other hand. The reason why the lingual configuration for /s#ʃ/ was /s/-like at onset and did not reach the /ʃ/ target until about the sequence midpoint could be sought in the fact that the preceding vowel was [ə] instead of [e] ([əs#ʃe] *tu compres xeixa* 'you buy candeal wheat'). It may be that, in contrast with /s#s/, which was also preceded by schwa in our database ([əs#se] *en compres cent* 'you buy one hundred of them'), during the onset of the /s#ʃ/ frication noise, while regressive assimilation occurs at constriction location, the tongue body behind constriction placement is still being raised and fronted from the appropriate lingual configuration for schwa to that for /ʃ/. Moreover, the reason why this change in tongue body position did not affect the acoustic spectrum must have been due to the fact that the COG frequency for lingual fricatives depends almost exclusively on the size of the cavity located in front of the constriction (Fant 1960). Thus, while /s#ʃ/ assimilates into [ʃ#ʃ] in Catalan, articulatory accommodation to the preceding vowel segment is allowed to occur at the back of the vocal tract as long as it does not jeopardize the frequency characteristics of the [ʃ] frication noise. This data interpretation finds support in the EPG data reported in Recasens & Mira (2013) showing the predicted assimilated outcome for /s#ʃ/ in sentences where the preceding vowel was mid front ([ɛs#ʃə] *es prenia tres xarops* '(s)he was taking three kinds of syrup', [es#ʃa] *li digué que matés xais* 'he asked him/her to kill some muttons').

The experimental findings reported in the present investigation are relevant with respect to more theoretical issues. Differences in articulatory adaptation between the two lingual fricatives reported in this and previous studies for various languages support the notion that /ʃ/ should be specified for a higher degree of articulatory constraint than /s/: /s/ assimilates to following /ʃ/ while /ʃ/ does not assimilate to following /s/; there may be carryover /ʃ/-to-/s/ effects which may be rendered phonological and thus lead to the assimilation of /s/ to preceding /ʃ/ (see Introduction). A higher degree of articulatory constraint for /ʃ/ than for /s/ appears to be due to a more precise tongue body positioning associated with the tongue raising gesture itself, the shaping of a relatively long laminal or lamino-predorsal constriction channel and the formation of a front sublingual cavity. Moreover, a rationale for why /s/ assimilates to following /ʃ/ rather than to preceding /ʃ/ and /ʃ#s/ is far more variable than /s#ʃ/ (the articulatory motion for /ʃ#s/ may show blending, progressive assimilation or two targets) should be sought in the different production strategies involved, i.e., free adaptation vs. repositioning (see the Introduction), and also in the fact

that regressive but not progressive assimilation conforms to phonemic preprogramming in speech, i.e., phonemes are programmed in advance of their actual articulatory manifestation. On the other hand, carryover coarticulation and progressive assimilation for /ʃ#s/ are likely to be directly related to the mechanico-inertial effects associated with the tongue dorsum raising/fronting gesture for the palatoalveolar fricative as well as for other dorsal consonants, as suggested by the realization [ɲʃ] and [ʎʃ] of the Catalan sequences /ɲs/ and /ʎs/ in words like *anys* 'year, pl.' and *alls* 'garlic, pl.'.

A cross-language comparison among the adaptation strategies occurring in the fricative consonant sequences under analysis reveals a general pattern towards the prevalence of /ʃ/ over /s/ at the regressive level rather than at the progressive level where the three adaptation strategies referred to above may be found and languages/dialects and speakers may differ regarding which one of these strategies is at work. Thus, acoustic data for four English subjects reported in Niebuhr et al. (2011) turned out to be very similar to those for the Eastern Catalan speakers reported in the present study with regressive assimilation occurring for /sʃ/ and a /ʃ/-to-/s/ trajectory for /ʃs/, which may proceed gradually or abruptly. EPG and acoustic data for English subjects provided in Pouplier et al. (2011) show a similar scenario with also instances of /ʃs/ where the [ʃ] realization extends until C2. Descriptive studies also refer to regressive assimilation for /sʃ/ in English (*this shop*; Cruttenden 2008: 302) and the EPG data presented in Nolan et al. (1996) are consistent with this segmental adaptation mechanism. In other languages/dialects, /ʃ/ is favored over /s/ both leftwards and rightwards. This appears to be the case in Southern French (Niebuhr et al. 2011), Western Catalan, and also European Portuguese where /s#ʃ/ is unavailable since word-final /s/ is realized as [ʃ] while /ʃ#s/ is implemented as [ʃ] ([dojˈʃaʃ] *dois chás* 'two teas', [dojʃɐˈpatuʃ] *dois sapatos* 'two shoes'; Leite de Vasconcelos 1901: 120; Mateus & d'Andrade 2000: 145). It seems therefore plausible to characterize as categorical regressive assimilation the adaptation scenario occurring in most instances of /sʃ/ and to allow for several production strategies for /ʃs/ that may depend on multiple factors, among which not only speaker and dialect but presumably speech rate and stress as well.

An issue open for further research is whether language-dependent differences in the extent to which /s/ assimilates to following /ʃ/ parallel a trend for categorical regressive place assimilations to be favored in certain languages rather than others where C1 may not assimilate to C2 or adapt to it partially in constriction location (i.e., gradient assimilation). This has been found to be the case for /n/+ velar, labial sequences in English and German which may exhibit assimilation, no assimilation or reduction of the alveolar gesture, as opposed to several Romance languages such as Italian, Spanish and Catalan where the nasal assimilates to the following heterorganic consonant practically all the time and traces of the C2 lingual gesture are not likely to occur during C1 (see Kochetov et al. 2021 for a summary and Bergman 2012 for articulatory data for /n#ɡ/ in German). Regarding the fricative consonant sequences of interest, the production of /s#ʃ/ has been shown to involve not only regressive assimilation but also blending in German (Pouplier & Hoole 2016), which clearly violates the prediction that /s/ should assimilate to /ʃ/ in these circumstances. Following a suggestion put forth in the Introduction, it may very well be that the reluctance for /s/ to assimilate to the following palatoalveolar fricative is due to speakers having laminodental productions of this consonant (and see in this respect Howson & Redford 2022 about the high degree of coarticulatory resistance for dental fricatives).

As also pointed out in the Introduction, we believe that the term 'dynamic blending' is suitable for cases of mutual adaptation in sequences of two consonants articulated with a central channel for the passage of airflow and thus lingual fricatives and lingual approximants (/ʃ#s/, /s#j/), a reason for this being that their articulatory trajectories are also dynamic when produced in VCV sequences. An issue for further research is how these blended realizations are perceived in running speech by listeners. Regarding Catalan, the

blended fricative outcome of /ʃ#s/ is often perceived as intermediate between /ʃ/ and /s/, i.e., as palatalized /s/ or depalatalized /ʃ/ (Badia 1951: 101), which is very much in accordance with the gestural blending account.

Generally speaking, a good match was obtained between the lingual and COG frequency trajectories, which is in agreement with that between the EPG and acoustic data reported in earlier studies (Nolan et al. 1996; Pouplier et al. 2011). A relevant issue is why the lingual profiles for /ʃ#s/ were more similar to those for the palatoalveolar fricative while the COG values for the same sequence approached the ones for the alveolar fricative instead. As pointed out in the Introduction, this may be related to the fact that, while the lingual profiles reflect changes in overall tongue configuration, the COG values for fricative consonants depend on the size of the cavity located in front of the lingual constriction. Regarding front cavity size, it needs to be recalled that /ʃ/ is produced not only with a more retracted constriction than /s/ but also with a larger sublingual cavity. In Catalan, the apical or laminal alveolar constriction for /s/ is rather back which means that it ought to be articulated with a larger sublingual space than the more anterior, predorsal variety of the alveolar fricative to be found in languages like French. This difference in constriction location causes the spectral peak of the frication noise to be relatively low, i.e., ca. 5000 Hz, when compared to that for /s/ in other languages (Jongman et al. 2000; Zygis 2003). In these circumstances, we speculate that a progressive reduction in front cavity size during the production of /ʃ#s/ as the lingual constriction is being fronted causes a fast spectral COG increase towards a value more appropriate for /s/. This change in constriction location could be related to a change in formant-cavity affiliation, i.e., the front cavity would be affiliated with F3 for /ʃ/ and with F3 or F4 for apicoalveolar /s/ instead of with F4 or F5, which would be the case if /s/ were predorso-dentoalveolar (Stevens 1989; Dart 1991; Tabain 2001).

In agreement with previous EPG and acoustic data for Catalan, the present study has confirmed a clear trend for /s/ to assimilate to following /ʃ/ and for the sequence /ʃ#s/ to exhibit a /ʃ/-to-/s/ trajectory, which at least for some speakers appears to result from a dynamic blending production strategy. Results may be accounted for assuming that the two consonants differ in degree of articulatory constraint and articulatory precision, as also revealed by a higher degree of coarticulatory resistance for /ʃ/ than for /s/, and that the C1-to-C2 transition for /ʃ#s/ involves articulatory repositioning and thus an asynchronous activation of different portions of the tongue. Moreover, the ultrasound data reported in the present investigation indicate that dynamic blending may be observed at specific regions of the tongue before others, i.e., at the tongue back before more anterior tongue regions. Future research could look into the language- and speaker-dependent factors which may intervene in the various realizations of /ʃs/ and how these realizations are perceived by listeners and also into deviations from the regressive assimilation pattern for /sʃ/ which may occur in languages other than German.

Appendix: Supplementary statistical data

Tables A1, A2, A3 and A4 provide estimate, standard error and p values for the post-hoc analysis results reported in Tables 1 through 4

Table A1. Estimate, standard error and p values for the post-hoc analysis results reported in the middle and bottom panels of Table 1 (XS=/ʃ#s/, SS=/s#s/, XX=/ʃ#ʃ/)

Zone	Temporal point	Sequence	Sequence	Estimate	Standard Error	p-value
ALV	1	XS	SS	3.0552	0.4016	<.0001
ALV	2	XS	SS	2.7940	0.4016	<.0001
ALV	3	XS	SS	1.8606	0.4016	<.0001
ALV	4	XS	SS	0.9994	0.4016	0.0141
ALV	5	XS	SS	0.2500	0.4016	0.0447
PAL	1	XS	SS	6.6744	0.4016	<.0001
PAL	2	XS	SS	6.0002	0.4016	<.0001
PAL	3	XS	SS	4.6462	0.4016	<.0001
PAL	4	XS	SS	2.6968	0.4016	<.0001
PAL	5	XS	SS	1.3050	0.4016	0.0015
VEL	1	XS	SS	−0.4434	0.4016	0.2716
VEL	2	XS	SS	−0.1144	0.4016	0.7762
VEL	3	XS	SS	0.0700	0.4016	0.8619
VEL	4	XS	SS	−0.1216	0.4016	0.7625
VEL	5	XS	SS	−0.4562	0.4016	0.2581
PHAR	1	XS	SS	−7.1896	0.4016	<.0001
PHAR	2	XS	SS	−6.9992	0.4016	<.0001
PHAR	3	XS	SS	−4.9934	0.4016	<.0001
PHAR	4	XS	SS	−2.8964	0.4016	<.0001
PHAR	5	XS	SS	−1.4890	0.4016	0.0003
ALV	1	XS	XX	−0.2454	0.4016	0.5422
ALV	2	XS	XX	−0.2986	0.4016	0.4585
ALV	3	XS	XX	−1.1328	0.4016	0.0056
ALV	4	XS	XX	−1.6706	0.4016	<.0001
ALV	5	XS	XX	−1.3196	0.4016	0.0013
PAL	1	XS	XX	−0.2912	0.4016	0.4697
PAL	2	XS	XX	−0.6704	0.4016	0.0975
PAL	3	XS	XX	−2.2062	0.4016	<.0001
PAL	4	XS	XX	−3.4558	0.4016	<.0001
PAL	5	XS	XX	−2.9254	0.4016	<.0001
VEL	1	XS	XX	0.4934	0.4016	0.2214
VEL	2	XS	XX	0.9772	0.4016	0.0163
VEL	3	XS	XX	0.9590	0.4016	0.0184
VEL	4	XS	XX	1.0714	0.4016	0.0086
VEL	5	XS	XX	1.0236	0.4016	0.0120
PHAR	1	XS	XX	2.6578	0.4016	<.0001
PHAR	2	XS	XX	2.3816	0.4016	<.0001
PHAR	3	XS	XX	3.1448	0.4016	<.0001
PHAR	4	XS	XX	3.6424	0.4016	<.0001
PHAR	5	XS	XX	2.6136	0.4016	<.0001

Table A1. Continued

Zone	Temporal point	Temporal point	Estimate	Standard Error	p-value
ALV	1	2	0.1590	0.4016	0.6928
ALV	1	3	1.1340	0.4016	0.0055
ALV	1	4	2.0492	0.4016	<.0001
ALV	1	5	2.6354	0.4016	<.0001
ALV	2	3	0.9750	0.4016	0.0166
ALV	2	4	1.8902	0.4016	<.0001
ALV	2	5	2.4764	0.4016	<.0001
ALV	3	4	0.9152	0.4016	0.0243
ALV	3	5	1.5014	0.4016	0.0003
ALV	4	5	0.5862	0.4016	0.1468
PAL	1	2	0.7476	0.4016	0.0649
PAL	1	3	2.4128	0.4016	<.0001
PAL	1	4	4.1008	0.4016	<.0001
PAL	1	5	4.1980	0.4016	<.0001
PAL	2	3	1.6652	0.4016	<.0001
PAL	2	4	3.3532	0.4016	<.0001
PAL	2	5	3.4504	0.4016	<.0001
PAL	3	4	1.6880	0.4016	<.0001
PAL	3	5	1.7852	0.4016	<.0001
PAL	4	5	0.0972	0.4016	0.8091
VEL	1	2	−0.0924	0.4016	0.8184
VEL	1	3	−0.0894	0.4016	0.8242
VEL	1	4	−0.0448	0.4016	0.9113
VEL	1	5	−0.1556	0.4016	0.6990
VEL	2	3	0.0030	0.4016	0.9941
VEL	2	4	0.0476	0.4016	0.9058
VEL	2	5	−0.0632	0.4016	0.8752
VEL	3	4	0.0446	0.4016	0.9117
VEL	3	5	−0.0662	0.4016	0.8693
VEL	4	5	−0.1108	0.4016	0.7831
PHAR	1	2	−0.3580	0.4016	0.3743
PHAR	1	3	−2.1864	0.4016	<.0001
PHAR	1	4	−3.4370	0.4016	<.0001
PHAR	1	5	−3.3312	0.4016	<.0001
PHAR	2	3	−1.8284	0.4016	<.0001
PHAR	2	4	−3.0790	0.4016	<.0001
PHAR	2	5	−2.9732	0.4016	<.0001
PHAR	3	4	−1.2506	0.4016	0.0023
PHAR	3	5	−1.1448	0.4016	0.0051
PHAR	4	5	0.1058	0.4016	0.7926

Table A2. Estimate, standard error and p values for the post-hoc analysis results reported in the middle and bottom panels of Table 2 (XS=/ʃ#s/, SS=/s#s/, XX=/ʃ#ʃ/)

Temporal Point	Sequence	Sequence	Estimate	Standard Error	p-value
1	XS	SS	−519.90	116.08	<.0001
2	XS	SS	−838.08	116.08	<.0001
3	XS	SS	−615.32	116.08	<.0001
4	XS	SS	−435.09	116.08	0.0007
5	XS	SS	−349.35	116.08	0.0051
1	XS	XX	170.86	116.08	0.1508
2	XS	XX	582.19	116.08	<.0001
3	XS	XX	998.17	116.08	<.0001
4	XS	XX	1305.09	116.08	<.0001
5	XS	XX	630.86	116.08	<.0001

Temporal point	Temporal point	Estimate	Standard Error	p-value
1	2	−441.18	116.08	0.0006
1	3	−792.32	116.08	<.0001
1	4	−940.59	116.08	<.0001
1	5	147.41	116.08	0.2133
2	3	−351.14	116.08	0.0049
2	4	−499.42	116.08	0.0001
2	5	588.59	116.08	<.0001
3	4	−148.27	116.08	0.2107
3	5	939.74	116.08	<.0001
4	5	1088.01	116.08	<.0001

Table A3. Estimate, standard error and p values for the post-hoc analysis results reported in the middle and bottom panels of Table 3 (SX=/s#ʃ/, SS=/s#s/, XX=/ʃ#ʃ/)

Zone	Temporal point	Sequence	Sequence	Estimate	Standard Error	p-value
ALV	1	SX	SS	1.3190	0.4481	0.0039
ALV	2	SX	SS	2.0244	0.4481	<.0001
ALV	3	SX	SS	2.9522	0.4481	<.0001
ALV	4	SX	SS	2.8354	0.4481	<.0001
ALV	5	SX	SS	1.6644	0.4481	0.0003
PAL	1	SX	SS	2.8338	0.4481	<.0001
PAL	2	SX	SS	4.1902	0.4481	<.0001
PAL	3	SX	SS	5.9544	0.4481	<.0001
PAL	4	SX	SS	5.7784	0.4481	<.0001
PAL	5	SX	SS	4.0520	0.4481	<.0001

Table A3. Continued

Zone	Temporal point	Sequence	Sequence	Estimate	Standard Error	p-value
VEL	1	SX	SS	−0.0864	0.4481	0.8474
VEL	2	SX	SS	−0.3324	0.4481	0.4595
VEL	3	SX	SS	−0.1522	0.4481	0.7347
VEL	4	SX	SS	−0.5408	0.4481	0.2297
VEL	5	SX	SS	−1.1128	0.4481	0.0143
PHAR	1	SX	SS	−1.7922	0.4481	0.0001
PHAR	2	SX	SS	−3.0246	0.4481	<.0001
PHAR	3	SX	SS	−4.3988	0.4481	<.0001
PHAR	4	SX	SS	−4.1532	0.4481	<.0001
PHAR	5	SX	SS	−2.7810	0.4481	<.0001
ALV	1	SX	XX	−1.9816	0.4481	<.0001
ALV	2	SX	XX	−1.0682	0.4481	0.0186
ALV	3	SX	XX	−0.0412	0.4481	0.9269
ALV	4	SX	XX	0.1654	0.4481	0.7126
ALV	5	SX	XX	0.0948	0.4481	0.8328
PAL	1	SX	XX	−4.1318	0.4481	<.0001
PAL	2	SX	XX	−2.4804	0.4481	<.0001
PAL	3	SX	XX	−0.8980	0.4481	0.0572
PAL	4	SX	XX	−0.3742	0.4481	0.4052
PAL	5	SX	XX	−0.1784	0.4481	0.6912
VEL	1	SX	XX	0.8504	0.4481	0.0600
VEL	2	SX	XX	0.7592	0.4481	0.0926
VEL	3	SX	XX	0.7368	0.4481	0.1026
VEL	4	SX	XX	0.6522	0.4481	0.1480
VEL	5	SX	XX	0.3670	0.4481	0.4143
PHAR	1	SX	XX	8.0552	0.4481	<.0001
PHAR	2	SX	XX	6.3562	0.4481	<.0001
PHAR	3	SX	XX	3.7394	0.4481	<.0001
PHAR	4	SX	XX	2.3856	0.4481	<.0001
PHAR	5	SX	XX	1.3216	0.4481	0.0038

Zone	Temporal point	Temporal point	Estimate	Standard Error	p-value
ALV	1	2	−0.8076	0.4481	0.0738
ALV	1	3	−1.6938	0.4481	0.0002
ALV	1	4	−1.5230	0.4481	0.0009
ALV	1	5	−0.5152	0.4481	0.2524
ALV	2	3	−0.8862	0.4481	0.0501
ALV	2	4	−0.7154	0.4481	0.1128

Table A3. Continued

Zone	Temporal point	Temporal point	Estimate	Standard Error	p-value
ALV	2	5	0.2924	0.4481	0.5152
ALV	3	4	0.1708	0.4481	0.7037
ALV	3	5	1.1786	0.4481	0.0096
ALV	4	5	1.0078	0.4481	0.0562
PAL	1	2	−1.2830	0.4481	0.0049
PAL	1	3	−2.7360	0.4481	<.0001
PAL	1	4	−2.8214	0.4481	<.0001
PAL	1	5	−2.3896	0.4481	<.0001
PAL	2	3	−1.4530	0.4481	0.0015
PAL	2	4	−1.5384	0.4481	0.0008
PAL	2	5	−1.1066	0.4481	0.0148
PAL	3	4	−0.0854	0.4481	0.8491
PAL	3	5	0.3464	0.4481	0.4409
PAL	4	5	0.4318	0.4481	0.3370
VEL	1	2	0.4826	0.4481	0.2835
VEL	1	3	0.4898	0.4481	0.2764
VEL	1	4	0.7314	0.4481	0.1051
VEL	1	5	0.8580	0.4481	0.0577
VEL	2	3	0.0072	0.4481	0.9872
VEL	2	4	0.2488	0.4481	0.5797
VEL	2	5	0.3754	0.4481	0.4037
VEL	3	4	0.2416	0.4481	0.5907
VEL	3	5	0.3682	0.4481	0.4127
VEL	4	5	0.1266	0.4481	0.7780
PHAR	1	2	1.0648	0.4481	0.0190
PHAR	1	3	2.6164	0.4481	<.0001
PHAR	1	4	3.2172	0.4481	<.0001
PHAR	1	5	3.3582	0.4481	<.0001
PHAR	2	3	1.5516	0.4481	0.0007
PHAR	2	4	2.1524	0.4481	<.0001
PHAR	2	5	2.2934	0.4481	<.0001
PHAR	3	4	0.6008	0.4481	0.1823
PHAR	3	5	0.7418	0.4481	0.1003
PHAR	4	5	0.1410	0.4481	0.7535

Table A4. Estimate, standard error and p values for the post-hoc analysis results reported in the middle and bottom panels of Table 4 (SX=/s#ʃ/, SS=/s#s/, XX=/ʃ#ʃ/)

Temporal Point	Sequence	Sequence	Estimate	Standard Error	p-value
1	SX	SS	−670.29	114.12	<.0001
2	SX	SS	−1272.38	114.12	<.0001
3	SX	SS	−1529.17	114.12	<.0001
4	SX	SS	−1737.87	114.12	<.0001
5	SX	SS	−899.21	114.12	<.0001
1	SX	XX	20.46	114.12	0.8588
2	SX	XX	147.89	114.12	0.2043
3	SX	XX	84.31	114.12	0.4654
4	SX	XX	2.31	114.12	0.9840
5	SX	XX	81.00	114.12	0.4830

Temporal Point	Temporal point	Estimate	Standard Error	p-value
1	2	−157.27	114.12	0.1777
1	3	−28.86	114.12	0.8020
1	4	211.79	114.12	0.0427
1	5	546.88	114.12	<.0001
2	3	128.41	114.12	0.2688
2	4	369.06	114.12	0.0028
2	5	704.15	114.12	<.0001
3	4	240.65	114.12	0.0429
3	5	575.74	114.12	<.0001
4	5	335.09	114.12	0.0061

References

Badia Margarit, Antonio. 1951. *Gramática histórica catalana*. Barcelona: Noguer.

Bergmann, Pia. 2012. Articulatory reduction and assimilation in n#g sequences in complex words in German. In Philip Hoole, Lasse Bombien, Marianne Pouplier, Christine Mooshammer & Barbara Kühnert (eds.), *Consonant clusters and structural complexity*, 311–344. Berlin & New York: Mouton de Gruyter.

Browman, Catherine P. & Goldstein, Louis. 1989. Articulatory gestures as phonological units. *Phonology*, 6, 201–251.

Browman, Catherine P. & Goldstein, Louis. 1990. Tiers in articulatory phonology, with some implications for casual speech. In John Kingston & Mary E. Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341–376. Cambridge: Cambridge University Press.

Browman, Catherine P. & Goldstein, Louis. 1992. Articulatory phonology: An overview. *Phonetica*, 49, 155–180.

Cho, Taehong, Jun, Sun-Ah & Ladefoged, Peter. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, 30, 193–228.

Cruttenden, Alan. 2008. *Gimson's pronunciation of English*, 7th ed. London: Hodder Education Publishers.

Dart, Sarah N. 1991. Articulatory and acoustic properties of apical and laminal articulations. *UCLA Working Papers in Phonetics*, 79.

Dart, Sarah N. 1998. Comparing French and English coronal consonant articulation. *Journal of Phonetics*, 26, 71–94.

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America*, 120, 407–415.

Fant, Gunnar. 1960. *Acoustic theory of speech production*. The Hague: Mouton.

Fitch, W. Tecumseh & Giedd, Jay. 1999. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America*, 106, 1511–1522.

Forrest, Karen, Weismer, Gary, Milenkovic, Paul H. & Dougall, Ronald N. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, 84, 115–123.

Frisch, Stefan A. & Wodzinski, Sylvie M. 2016. Velar–vowel coarticulation in a virtual target model of stop production. *Journal of Phonetics*, 56, 52–65.

Guzik, Karita M. & Harrington, Jonathan. (2007) The quantification of place of articulation assimilation in electropalatographic data using the similarity index (SI). *Advances in Speech Language Pathology*, 9, 109–119.

Holst, Tara & Nolan, Francis. 1995. The influence of syntactic structure on [s] to [ʃ] assimilation. In Bruce Connell & Amalia Arvaniti (eds.), *Phonology and phonetic evidence. Papers in laboratory phonology IV*, 315–333. Cambridge: Cambridge University Press.

Howson, Phil J. & Redford, Melissa. 2022. A cross-sectional age group study of coarticulatory resistance: The case of late-acquired voiceless fricatives in English. *Journal of Speech and Hearing Research*, 65, 3316–3336.

Iskarous, K., Shadle, C. H. & Proctor, M. I. 2011. Articulatory–acoustic kinematics: The production of American English /s/. *The Journal of the Acoustical Society of America*, 129, 944–954.

Jongman, A., Wayland, R. & Wong, S. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, 108, 1252–1263.

Kochetov, Alexei, Colantoni, Laura & Steele, Jeffrey. 2021. Variable assimilation of English word-final /n/: Electropalatographic evidence. *English Language & Linguistics*, 25, 687–718.

Ladefoged, Peter & Maddieson, Ian. 1996. *The sounds of the world's languages.* Oxford/Cambridge, MA: Blackwell.

Leite de Vasconcelos, José. 1901. *Esquisse d'une dialectologie portugaise*. Paris: Aillaud.

Mateus, Maria Helena & d'Andrade, Ernesto. 2000. *The phonology of Portuguese*. Oxford: Oxford University Press.

Mielke, Jeff. 2015. An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *Journal of the Acoustical Society of America*, 137, 2858–2869.

Navarro Tomás, Tomás. 1918. *Manual de pronunciación española.* Madrid: Consejo Superior de Investigaciones Centíficas.

Niebuhr, Oliver, Clayards, Meghan, Meunier, Christine & Lancia, Leonardo. 2011. On place assimilation in sibilant sequences—Comparing French and English. *Journal of Phonetics*, 39, 429–441.

Nolan, Francis, Holst, Tara & Kühnert, Barbara. 1996. Modelling [s] to [ʃ] accommodation in English. *Journal of Phonetics*, 24, 113–137.

Perkell, Joseph S., Boyce, Suzanne E. & Stevens, Kenneth N. 1979. Articulatory and acoustic correlates of the [s-ʃ] distinction. In J. J. Wolf & Dennis H. Klatt (eds.), *97th Meeting of the Acoustical Society of America*, 109–113. New York: American Institute of Physics.

Perkell, Joseph S., Guenther, Frank H., Lane, Harlan, Marrone, N., Matthies, Melanie L., Stockmann, Ellen, Tiede, Mark & Zandipour, Majid. 2013. Production and perception of phoneme contrasts covary across speakers. In Jonathan Harrington & Marija Tabain (eds.), *Speech production: Models, phonetic processes, and techniques*, 69–84. New York: Psychology Press.

Pouplier, Marianne & Hoole, Philip. 2016. Articulatory and acoustic characteristics of German fricative clusters. *Phonetica*, 73, 52–78.

Pouplier, Marianne, Hoole, Philip & Scobbie, James M. 2011. Investigating the asymmetry of English sibilant assimilation: Acoustic and EPG data. *Laboratory Phonology*, 2, 1–33.

R Development Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Recasens, Daniel. 2006. Integrating coarticulation, assimilation and blending into a model of articulatory constraints. In Louis Goldstein, Douglas H. Whalen & Catherine T. Best (eds.), *Laboratory phonology 8*, 611–634. Berlin/New York: Mouton de Gruyter.

Recasens, Daniel. 2014. *Fonètica i fonologia experimentals del català. Vocals i consonants.* Barcelona: Institut d'Estudis Catalans.

Recasens, Daniel & Espinosa, Aina. 2009. An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America*, 125, 2288–2298.

Recasens, Daniel & Mira, Meritxell. 2013. An articulatory and acoustic study of the fricative clusters /sʃ/ and /ʃs/ in Catalan. *Phonetica*, 70, 298–322.

Recasens, Daniel & Rodríguez, Clara. 2017. Contextual and syllabic effects in heterosyllabic consonant sequences. An ultrasound study. *Speech Communication*, 96, 150–167.

Recasens, Daniel, Fontdevila, Jordi & Pallarés, Maria Dolors. 1993. An electropalatographic study of stop consonant clusters. *Speech Communication*, 12, 335–355.

Reidy, Patrick F. 2015. *The spectral dynamics of voiceless sibilant fricatives in English and Japanese*. Ph.D. dissertation, The Ohio State University.

Stevens, Kenneth N. 1989. On the quantal nature of speech. *Journal of Phonetics*, 17, 3–45.

Stone, Maureen. 2005. A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics*, 19, 455–501.

Tabain, Marija. 2001. Variability in fricative production and spectra: Implications for the hyper- and hypo-quantal theories of speech production. *Language and Speech*, 44, 57–94.

Zsiga, Elizabeth C. 1995. An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. In Bruce Connell & Amalia Arvaniti (eds.), *Phonology and phonetic evidence. Papers in laboratory phonology IV*, 282–302. Cambridge: Cambridge University Press.

Zygis, Marzena. 2003. Phonetic and phonological aspects of Slavic sibilant fricatives. *ZAS Papers in Linguistics*, 32, 175–213.